THE EFFECTS OF LAND COVER ON THE SPATIAL DISTRIBUTION OF LYME DISEASE IN NORTHERN VIRGINIA SINCE 2005

MEGAN NICOLE STEVENSON

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
GEOGRAPHY

Korine Kolivras, Committee Chair

Yang Shao

Cassidy Rist

September 25th, 2019

Blacksburg, Virginia

Keywords:

Lyme disease, Medical Geography, Land Cover, GIS, ZI Gamma Distribution, Forest Fragmentation, Virginia

THE EFFECTS OF LAND COVER ON THE SPATIAL DISTRIBUTION OF LYME DISEASE IN NORTHERN VIRGINIA SINCE 2005

Megan Stevenson

ABSTRACT

Lyme disease's spatial distribution and total number of cases have been increasing for several decades since it was first identified in the 1970s in New England. As a result, Lyme disease is the most common vector-borne disease in the United States. There have been various studies that have analyzed the factors that contribute to the expansion of Lyme disease's range, but few have looked specifically at how the land cover in the northern Virginia region has impacted the spatial distribution and emergence of the disease in the region. Therefore, this research analyzes the effect that land use/land cover (LULC) has on the spatial distribution of Lyme disease in northern Virginia from 2005 to 2017.

Human Lyme disease case data was received from the Virginia Department of Health (VDH) and 30 meter resolution National Land Cover Dataset (NLCD) images for the years 2001, 2006, and 2011 were gathered from the Multi-Resolution Land Characteristics Consortium's (MRLC) resources. The human Lyme disease case data were aggregated to their corresponding census tract and the incidence rate for each census tract for each year of the study was calculated based on annual population totals. The original land cover classifications were simplified to be comprised of four primary land cover types: i. Water, ii. Developed, iii. Forest, and iv. Herbaceous. The Geospatial Modelling Environment software was used in conjunction with ArcGIS Pro to collect information about the land cover characteristics. Variables were then derived for the statistical analysis, Generalized Regression with Zero Inflation Gamma Distribution.

The results of this study are in agreement with studies of a similar scope. The statistical tests suggest that there is a significant relationship between the incidence rate of human Lyme disease and various land cover types and edges. Primarily, the land cover types and edges that tend to have a significant relationship with incidence rate are developed land, forested land, herbaceous land, developed/forested edge, developed/herbaceous edge, and forested/herbaceous edge. While the LULC did not drastically change during the study's timeframe, it is still evident that the total number of human Lyme disease cases fluctuates as the LULC in the study area is altered.

The overall goal of this research is for the findings to be useful for areas that have similar land cover trends to that of the northern Virginia study area from this project. The results of this study confirm findings from past research that the way land is used and developed has an impact on the spatial distribution of Lyme disease. Careful consideration of the way land is altered could greatly increase public health safety in areas that are at risk of Lyme disease emergence.

THE EFFECTS OF LAND COVER ON THE SPATIAL DISTRIBUTION OF LYME DISEASE IN NORTHERN VIRGINIA SINCE 2005

Megan Stevenson

GENERAL AUDIENCE ABSTRACT

Lyme disease has been a growing problem in the United States over the last few decades, and is currently the most common vector-borne disease in the country. This research evaluates the land cover within specified counties of northern Virginia to determine if a correlation exists between forest fragmentation, suburbanization, and cases of human Lyme disease as has been demonstrated in other Lyme endemic regions in the United States. Few studies have focused specifically on northern Virginia when considering the impacts of land cover change on Lyme disease. Discovered through the use of geospatial and statistical analysis, the cluster of Lyme disease cases in northern Virginia are associated with forest fragmentation within the study region, which creates an ideal habitat for black-legged ticks and the white-footed mouse, allowing for an increase in Lyme disease transfer from vector to humans. The goal is for the research findings to be applicable to other regions with similar land cover types. Regions with similar characteristics would then be able to recognize the potential risk of human Lyme disease and implement ways to reduce the Lyme disease risk associated with suburban development.

The purpose of this study is to answer the following research questions: 1) How has the spatial distribution of Lyme disease in Northern Virginia changed since 2005 with respect to land cover? 2) Which suburban communities are more at risk for Lyme disease when considering their land cover types and the increasing spatial distribution of Lyme disease?

ACKNOWLEDGEMENTS

I would like to first thank my advisor and committee chair, Dr. Korine Kolivras. I appreciate all of the guidance, support, and patience that you provided me the last couple of years. Without your encouragement to pursue my Masters and your generosity in answering my many questions prior to applying to grad school, I'm not sure where I would be. I cannot thank you enough for pushing me to follow my interests in medical geography and GIS. I would also like to thank my committee, Dr. Yang Shao and Dr. Cassidy Rist. I appreciate your insights and recommendations throughout this process.

Next, I would like to thank the Virginia Department of Health (VDH) for supplying me with the human Lyme disease case data and their willingness to clarify any questions I had. I would also like to thank Liz Suekep, whose research was the basis for my own methodology. The work that Liz completed contributed to our field in a substantial way and created a standard for future research to follow. I also appreciate the help and guidance from my fellow research group members, Dr. Pyrros Alexander Telionis and Dr. Elizabeth Weaver. You both have impressive skills and experience within our field and beyond. Alex, thank you for recommending the land cover analysis software – it saved me a lot of time and was an extremely useful tool. Liz, thank you for assisting me with the statistical analysis and sharing your knowledge of PCA – it was incredibly helpful. Along those lines, I want to thank the Statistical Applications and Innovations Group (SAIG) at Virginia Tech for providing their assistance throughout the process of selecting and running statistical tests for my research.

I would also like to thank the faculty of the Department of Geography at Virginia Tech. You all peaked my interest in the geography field and served as wonderful mentors throughout my collegiate career.

Lastly, I want to thank my family and friends for being so supportive throughout this entire process. To my parents, thank you for always believing in me and encouraging me to achieve my goals. Your love and support helped to shape who I am today and I truly appreciate all of the opportunities you have provided me with and encouraged me to pursue. To my sister, thank you for being my best friend and never failing to be simply one phone call away. You were my first friend at Virginia Tech and have always been my role model. Thanks for showing me what it means to work hard and chase your dreams. To my friends, thank you for supporting me and reminding me to take breaks every now and then to enjoy the moment.

ATTRIBUTION

Dr. Korine Kolivras is my academic advisor and served as the committee chair. Her in-depth knowledge of both medical geography and Lyme disease were extremely helpful throughout the entirety of this project. She is very well-versed on the subject matter and provided me with great insights and guidance while developing my research questions and methods, analyzing results, and writing the comprehensive manuscript.

Dr. Yang Shao is a committee member from the Department of Geography. His experience with geospatial analysis was appreciated during the stage of research development, analysis, and writing.

Dr. Cassidy Rist is a committee member from the Department of Population Health Sciences. Her background and knowledge of Lyme disease were helpful throughout the research and writing processes.

TABLE OF CONTENTS

CHAPTER 1: PROBLEM STATEMENT

1.1 Introduction

Lyme disease has been an increasing problem in the eastern United States for the last several decades. According to the Centers for Disease Control and Prevention (CDC), approximately 95% of cases in 2015 were reported from fourteen states, most of which are on the east coast, with the two outliers being Minnesota and Wisconsin (CDC, 2017b). The increase in the spatial distribution of the disease occurred concurrently with the occurrence of a steady rise in the number of cases. The CDC (2017a) reports that in 1996 there were approximately 15,000 reported cases annually, and by 2016 that number had increased to nearly 35,000 reported cases a year (Figure 1.1). The number of reported cases is dwarfed by the estimated 300,000 cases of Lyme disease that likely occur within a year since many go unreported (CDC, 2017a, CDC, 2017b). When assuming that there are roughly 300,000 cases a year, it is estimated that the process of diagnosis and treatment for Lyme disease in the United States cumulatively costs upwards of $1 billion every year (Adrion et al., 2015).

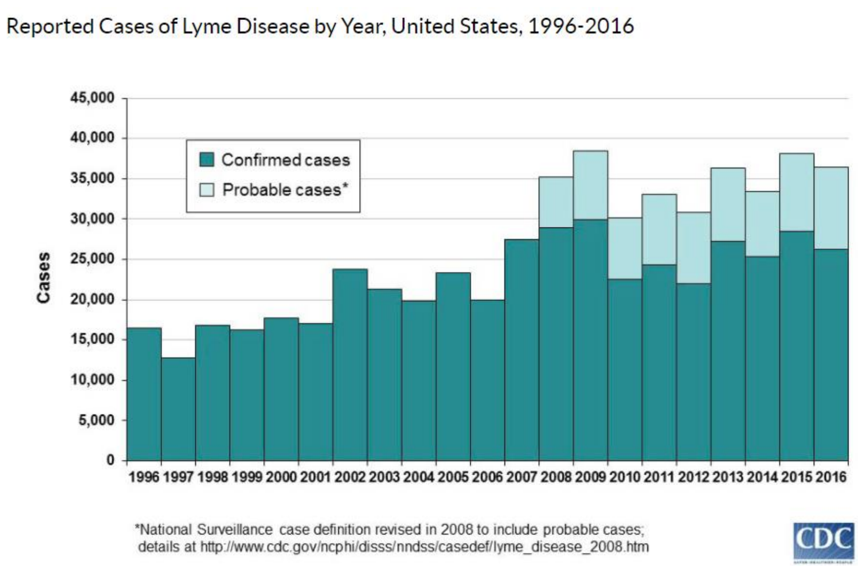

Figure 1.1: Graph of Lyme disease cases (1996-2016) in the United States per year.

1.2 Lyme disease at the State Level

Virginia, in particular, has seen an increase in the spatial distribution and total number of Lyme disease cases since 2005. The CDC reports that in 2006, Virginia had 357 reported cases of Lyme disease, and by 2015 Virginia had 1,102 confirmed cases (CDC, 2017a). The early cases of Lyme disease in Virginia were mostly concentrated in the northern portion of the state, but the disease has since spread to the southwestern region where a high number of cases has been reported in recent years (Brinkerhoff et al., 2014, Li et al., 2014). This study is focused on the northern Virginia region, which includes the nine counties and four independent cities seen in Figure 1.2. These counties are fairly urbanized, but also have large expanses of farmland and forestland, which makes the region an interesting place to examine Lyme disease as previous research has noted the importance of varied land cover and forest fragmentation in the disease transmission cycle. While a number of studies have examined links between land cover and Lyme disease, little research has focused on analyzing the effects of land cover and suburban sprawl on the spatial distribution and emergence of Lyme disease in northern Virginia. The

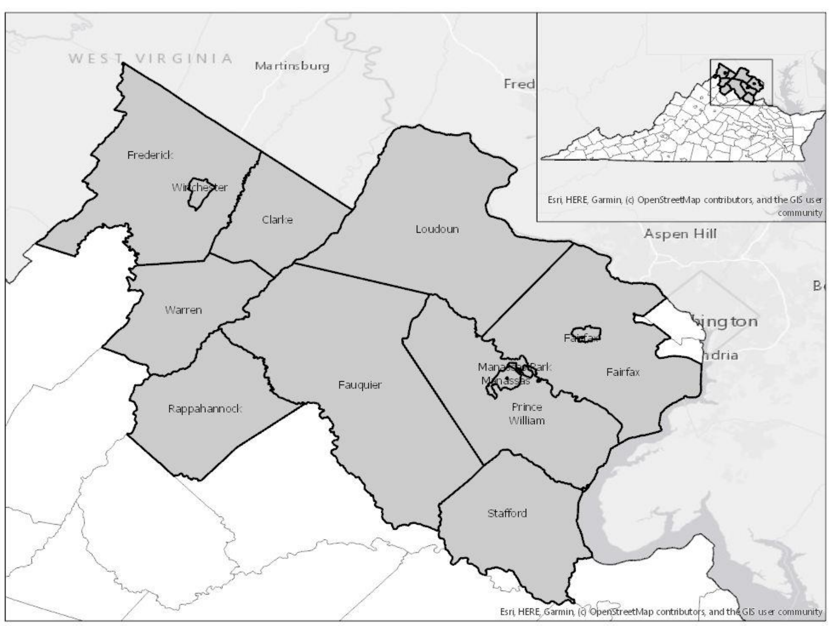

Figure 1.2: Defined study area includes: Clarke, Fairfax, Fauquier, Frederick, Loudoun, Prince William, Rappahannock, Stafford, and Warren counties, and Fairfax, Manassas, Manassas Park, and Winchester independent cities. County and state data layers from census.gov. Basemap from ESRI ArcPro software.

conclusions drawn from this study aim to help raise awareness of potential at-risk suburban communities and encourage the development of suburban areas in ways that will lower the threat of Lyme disease transmission to humans.

1.3 Problem Statement

By increasing our understanding of the underlying ecological processes that affect the spatial distribution of Lyme disease in an urban and semi-urban area like northern Virginia, we can apply that understanding to other regions with similar environmental and demographic characteristics. In turn, we will have a better idea of which areas are most at risk of seeing an increase in the number of Lyme disease cases elsewhere, and can design future developed areas in a manner that will reduce the likelihood of Lyme disease transmission. In addition, with the knowledge of how the spatial distribution of Lyme disease is related to land cover, people can make more educated choices and public health officials can be more aware of potential cases. The hope is that the conclusions drawn from this study will provide useful information on how to reduce the risk of human Lyme disease in developed and semi-developed areas.

As described previously, there are basic and applied research needs related to Lyme disease in northern Virginia. I will focus on the following research questions:

– What land cover characteristics are related to the emergence of Lyme disease in Northern Virginia from 2005 through 2017?

– Where should the Virginia Department of Health concentrate Lyme disease education and surveillance efforts when considering land cover types and the increasing spatial distribution of Lyme disease?

References

Adrion, E. R., et al. (2015). "Health Care Costs, Utilization and Patterns of Care following Lyme Disease." PLOS ONE 10(2): e0116767.

Brinkerhoff, R. J., et al. (2014). "Lyme disease, Virginia, USA, 2000-2011." Emerging infectious diseases 20(10): 1661.

Centers for Disease Control and Prevention (n.d.). *Lyme Disease (Borrelia burgdorferi) 2008 Case Definition*. Retrieved from https://wwwn.cdc.gov/nndss/conditions/lyme-disease/case-definition/2008/

Centers for Disease Control and Prevention. (2017a). *Data and Statistics*. Retrieved from https://www.cdc.gov/lyme/stats/index.html

Centers for Disease Control and Prevention. (2017b). *Why is CDC concerned about Lyme disease?* Retrieved from https://www.cdc.gov/lyme/why-is-cdc-concerned-about-lyme-disease.html

Figures Reference List

Centers for Disease Control and Prevention. (2017). *Lyme disease graphs: Reported cases of Lyme disease by year, 1996-2016*. Retrieved from https://www.cdc.gov/lyme/stats/graphs.html

Esri. (2018). ArcGIS Pro: Release 2.2.1 [Computer software]. Retrieved from https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview

United States Census Bureau, Geography Division. (2017). 2017 TIGER/Line Shapefiles. Retrieved from https://www.census.gov/cgi-bin/geo/shapefiles/index.php

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter discusses the findings and methods of other works that relate to the research questions of this project. The literature review will include sections about the sub-discipline in which this study falls under, Lyme disease on large and small scales, and the effects of forest fragmentation. Each of these sections carries an importance in the overall understanding of Lyme disease and its prevalence within the study area.

2.2 Medical Geography

This study is grounded in the sub-discipline of medical geography, which is the study of the spatial distribution of health, disease, and access to healthcare. Hippocrates is often credited for first introducing the idea of what has become medical geography by suggesting the use of a holistic approach that recognizes the importance of both culture and the environment in examining the health of a population (Meade, 2014). The discipline can be used to analyze health and disease patterns from the past and present, as well as to predict future occurrences. A key aspect of medical geography involves understanding previous distributions of disease and the environments that are most prone to an outbreak. Thus, through the use of theories and techniques in medical geography, predictions of at-risk areas can be derived, which allows for appropriate procedures to be put in place with the hopes of reducing the impacts of future disease events. The information gathered through these processes can lead to an understanding of why certain communities have higher levels of illness and death so that public health officials can implement solutions. Overall, medical geography encompasses the relationships between environmental, social, and economic aspects and how these unique relationships effect the health of an area. This research will contribute to expanding the understanding of vector-borne disease patterns, specifically Lyme disease, and the impact that land cover has on our ability to predict and reduce increasing trends of disease.

2.2.1 Landscape Epidemiology

This research is supported by landscape epidemiology, a theoretical framework in medical geography that can be described as understanding the relationships between "landscape heterogeneity and the underlying ecological processes that drive the spread and persistence of disease" (Meentemeyer, Haas, & Vaclavik, 2012). More specifically, these relationships involve the host, vector, and pathogen of a disease within an environment that foster either emergence or the potential for increased prevalence of the disease in a region (Reisen, 2009). The way landscape factors interact with surrounding environmental and demographic dynamics can determine if a disease will thrive in that area. The spatial element of disease transmission is another important factor involved in landscape epidemiology (Reisen, 2009). Lambin et al. (2010) found that three factors come into play when looking at the spatial distribution of a disease: "(i) the pathogenic cycle and the biology of vectors, hosts and pathogens; (ii) ecosystem processes at the landscape scale, as influenced by ecosystem structure and composition, landscape connectivity and configuration, climate, species interactions; and (iii) land use, human behavior and mobility, knowledge and perception of disease risk, and socio-economic conditions" (pg 11).

Many different tools that can be used to analyze and predict spatial patterns of disease including geographic information systems (GIS), remote sensing, and global positioning systems (GPS). In terms of medical geography, GIS can be used to overlay multiple data sets to visualize trends and key factors that regions with disease outbreaks have in common (Kitron, 1998). As shown by Hendricks & Mark-Carew (2017), cluster analysis is an effective way to track patterns and disease hot spots, where hot spots can be defined as an area with a high number of reported cases of a disease. This method of cluster analysis can be combined with GIS to provide another method of creating visual representations of a disease on a map. GIS can also be used to delineate the spatial distributions of various populations involved in a disease study, such as the reservoir/host, vector, and humans (Kitron, 1998). Previous studies have used GIS to track vector-borne diseases, like Lyme disease and malaria, and to identify various risk factors by using the statistical capabilities of the software combined with the visual display of the data on a map (Kitron, 1998). Similarly, remote sensing is useful in identifying habitats, tracking changes in land cover, and predicting shifts in disease patterns due to the relationship between land cover and vector habitat, as suggested by Kitron (1998). Brownstein et al (2005) used remote sensing to measure the normalized difference vegetation index (NDVI) on satellite images to classify land cover and identify forested areas, so that black-legged tick collection sampling sites could be better selected. Both GIS and remote sensing can be combined with data collected by GPS for further analysis of data and to formulate conclusions or predictions about the spatial patterns. For example, when examining vector-borne diseases, GPS can be useful when tracking specific populations of animals, or for the purpose of mapping exact locations by overlaying the data points on GIS layers or remotely sensed images.

2.3 Lyme disease in the United States

Lyme disease originated in Lyme, Connecticut, and is the most common vector-borne disease among humans in the United States (Brownstein et al., 2005, Cromley et al., 1998, Guerra et al., 2002). The black-legged tick, also known as *Ixodes scapularis,* is the east coast vector of the causal agent of Lyme disease, the bacterium *Borrelia burgdorferi* (CDC, 2019).

The most competent reservoir of the Lyme disease bacterium on the east coast is the white-footed mouse (Ostfeld, 2011). Therefore, black-legged tick larvae that feed on an infected white-footed mouse might acquire the bacterium and potentially transmit it to humans later in their life cycle (Allan et al., 2002). The common belief is that black-legged ticks are the most likely to transmit Lyme disease to humans when they are in the nymph stage due to their small size, which makes them hard to see, and the fact that the peak nymph season for black-legged ticks is summer when people are more likely to be outside (Allan et al., 2002, Brinkerhoff et al., 2014, Guerra et al., 2002, Johnson et al., 2017); however, transmission of the bacterium can occur at the adult stage as well. After a bite from an infected tick, symptoms may appear. The easiest way to identify a Lyme disease infection is by the bullseye rash, otherwise known as erythema migrans, which appears at the site of the tick bite for roughly 70-80% of infected people (CDC, 2016b). Other common symptoms include headache, flu-like symptoms, joint pain, and fatigue.

Lyme disease was first identified in the mid-1970s (Steere et al., 2004), and cases have been reported to and tracked by the CDC in the United States since 1991 (Kugeler et al., 2015). Hawaii is the only state in the U.S. with zero reported cases of human Lyme disease. In some states where human Lyme disease incidence is low, the cases may be attributed to travel (Hendricks & Mark-Carew, 2017). Cases of human Lyme disease are most prevalent in the eastern and mid-Atlantic U.S., but a large number of cases have also been reported in several Midwestern states, namely Minnesota and Wisconsin (CDC, 2017b, Kugeler et al., 2015). From 1993 to 2012, the number of counties with high incidence rates of Lyme disease within previously identified high incidence regions increased greatly (Kugeler et al., 2015).

In addition to increased incidence rates, the range of Lyme disease has expanded. The expansion of Lyme disease cases in the northeast epicenter typically moved in the westward and northward directions (Kugeler et al., 2015). Others have noted the expansion of Lyme disease in a southward direction away from the northeast US epicenter, specifically towards the southwestern region of Virginia (Brinkerhoff et al, 2014, Lantos et al., 2015, Li et al., 2014). Hendricks and Mark-Carew (2017) examined reported cases of human Lyme disease in Kentucky, Maryland, Ohio, Pennsylvania, Virginia, and West Virginia and observed similar expansion and reporting trends as those characteristic in the national expansion trends. There seemed to be higher levels of reporting in the eastern portion of the study area, which Hendricks and Mark-Carew (2017) believed to be associated with a greater understanding and awareness of Lyme disease presence among citizens and healthcare providers. The results of their study suggested a link between human behavior and environmental surroundings as a key component of the spread of human Lyme disease (Hendricks & Mark-Carew, 2017). Rates of human Lyme disease have been studied in eastern United States national parks for the intention of informing visitors of the risks and raising overall awareness (Johnson et al., 2017). The study by Johnson et al. (2017) evaluated risk within nine national parks and the findings were consistent with areas surrounding the national parks. Johnson et al. (2017) concluded that the spatial distribution of black-legged ticks carrying the bacterium was expanding westward, which matches the nationally observed trends. Since the early 1990's, Lyme disease in Wisconsin has expanded from the northwest region of the state towards the south and the east due to the increasing range of the black-legged tick (CDC, 2016a, Guerra et al., 2002). There was a large increase in the black-legged tick population in the southwestern portion of the state, in part because the region has ideal habitat conditions (Guerra et al., 2002). Similar to findings in other states, human Lyme disease cases were found to correlate with ideal black-legged tick habitat in the surrounding area (Guerra et al., 2002).

The areas where Lyme disease occurs at the highest levels typically have well-established populations of black-legged ticks, as well as white-footed mice and white-tailed deer, both of which are known to be a common source of blood meals for black-legged ticks (although deer are not a reservoir, or source of the bacterium) (e.g., Allan et al., 2002, Brinkerhoff et al., 2014, Brownstein et al., 2005, Guerra et al., 2002). In terms of ticks, a population is considered to be well-established if groupings of each of the three stages of the tick's life cycle are present (Guerra et al., 2002).

2.4 Lyme disease in Virginia

The case total and spatial distribution of Lyme disease in Virginia has increased from 2000 to their highest level at the present. Between 2000 and 2006, there were fewer than 400 reported cases of Lyme disease in Virginia per year (Lantos et al., 2015), but by 2014, the annual number of cases had increased to 1,346 (Lantos et al., 2015). From 2001 to 2011, there is an evident trend of the disease expanding across the state in the southwestern direction (Brinkerhoff et al, 2014, Lantos et al., 2015, Li et al., 2014). The years 2000 to 2005 are considered to be the early stages of the disease's presence in Virginia. During this time, the eastern shore region and the northern region of Virginia were hot spots for Lyme disease (Brinkerhoff et al., 2014). After 2005, large clusters of Lyme disease cases developed in the southwestern region of Virginia with Roanoke, Lynchburg, and Blacksburg highlighted as the areas with the highest densities of reported Lyme disease cases in the state (Lantos et al, 2015). When looking at human Lyme disease incidence in the northern and eastern regions of Virginia, most of the cases were found in communities along major roadways, such as highways I-81, I-95, and I-64 (Li et al., 2014). These well-populated areas typically have suburban communities that create ideal habitat conditions for both the black-legged tick and the transmission of Lyme disease to humans (Li et al., 2014). Lantos et al. (2015) states that trends of human Lyme disease cases in Virginia were the clearest in the northern portion of the state, which is more populated and more geographically similar to northern states with high incidence rates. This finding further suggests that areas of suburbanization foster the ideal habitat for Lyme disease transmission and supports the need for this study on Lyme disease in northern Virginia.

When considering the spread and emergence of Lyme disease, important information can be drawn from clusters, such as in which direction and at what speed is Lyme disease spreading (Li et al., 2014). Hendricks & Mark-Carew (2017) used spatial weights by year as a clustering method to identify Lyme disease hot spots and analyze changes or new areas of emergence. Similarly, a study conducted by Lantos et al. (2015) used spatial cluster analysis to identify the major areas of Lyme disease cases within the desired study region. Along with spatial clustering, Lantos et al. (2015) also performed spatiotemporal cluster analysis as a way of comparison between time periods. Through the combination of spatial and spatiotemporal cluster analyses, Lantos et al. (2015) demonstrated that in 2000 there was a cluster of Lyme disease cases in northern Virginia, but post-2000 it expanded in a southwest direction.

2.5 Effects of Land Cover and Forest Fragmentation on Lyme disease

In most cases, the emergence and spread of Lyme disease is related to changes in land cover, specifically a shift to a suburbanized development (Li et al., 2014). Shortly after European colonization of the northeastern US, deforestation took place in order to clear land for agricultural and residential purposes. Eventually more of the land was converted to residential use and fragments of previously forested land grew back in the form of second growth forests (Mayer, 2000). As the northeastern U.S. experienced increasing suburban residential development, the remaining original forested land was further cleared and fragmented, and as a result, cases of human Lyme disease increased soon after due to the creation of ideal habitat for both the white-footed mouse, which is the Lyme disease reservoir, and the white-tailed deer, which is the travel agent of the black-legged tick (Cromley et al., 1998, Dymond et al., 2013, Li et al., 2014, Mayer, 2000).

Forest fragmentation is the division of forested land into smaller segments. The sections of fragmented forest are bordered by other land cover types such as lawn, shrub land, agricultural fields, developed areas, and other land cover. A driving factor in the spread of human Lyme disease is forest fragmentation, especially when paired with suburbanization (Brownstein et al., 2005, Cromley et al., 1998, Jackson et al., 2006). As mentioned in the previous section, areas in both northern and eastern Virginia had high incidence of human Lyme disease along highways partly due to the fact that they are areas with increasing population growth and suburbanization (Li et al., 2014).

A change in land cover from intact forests to forest fragments in suburban neighborhoods creates habitats well-suited for the white-footed mouse and the white-tailed deer, which subsequently creates an ideal habitat for the Lyme disease transmission cycle. Areas with fragmented forest edges tend to have an adequate availability of hosts for the black-legged tick to receive blood meals (Brownstein et al., 2005, Jackson et al., 2006, Suekep et al., 2015). The forest edge allows for white-footed mice and white-tailed deer to graze in the herbaceous land cover and retreat to the forest patches for safety (Seukep et al., 2015). There is also a decreased threat to the white-footed mouse and the white-tailed deer populations since their common predators do not typically live in areas with limited forested land as biodiversity is generally lower in forest fragments (Allan et al., 2003, Brownstein et al., 2005). Due to the heightened availability of white-footed mice in areas of forest fragmentation, black-legged ticks are more likely to receive blood meals from the mice, thus increasing the likelihood of acquiring the Lyme disease bacterium (Allan et al., 2003); infected white-footed mice are highly competent reservoirs, transmitting the bacterium to 90% of the ticks that feed upon them (Ostfeld, 2011). The movement of white-footed mice and the white-tailed deer between the forest and herbaceous edge increases the mobility of the black-legged tick, which increases the potential of its interaction with humans. Humans could easily be exposed to black-legged ticks while in their yards, walking on nearby paths, playing in a field, or from interacting with their pets (Finch et al., 2014, Hendricks & Mark-Carew, 2017).

Specific landscape configuration can affect Lyme disease transmission. Jackson et al. (2006) explains that the way in which the forest is fragmented plays a role in the risk of Lyme disease as well. The more divided a forest is, the smaller the forest sections will be and more forest-herbaceous edges will be exposed. An area with these conditions is more likely to have a Lyme disease threat than an area with larger forest patches and very few forest edges (Brownstein et al., 2005, Jackson et al., 2006). Jackson et al. (2006) explains how two areas could be equally split with 50% forest and 50% herbaceous land cover, but if one has more forest-herbaceous edges it will have a higher risk of human Lyme disease incidence (Figure 2.1).

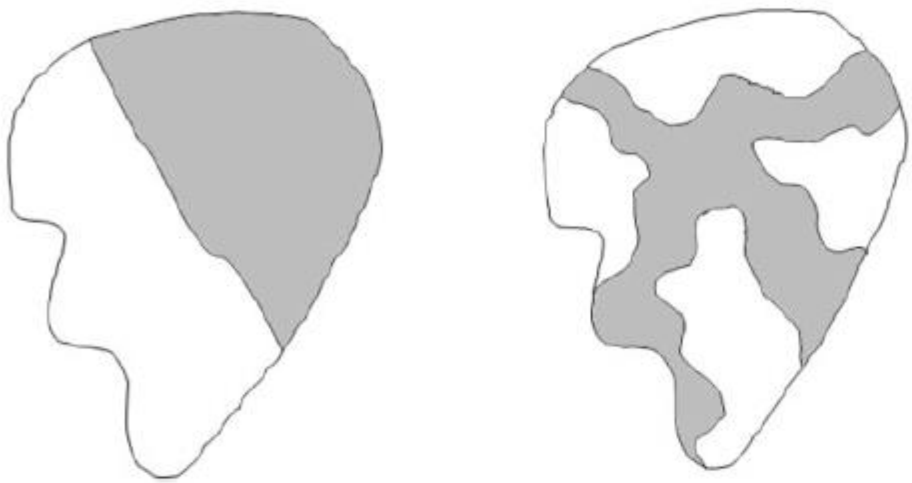

Figure 2.1: The graphics above both show 50% forest cover (dark grey shading). The graphic on the left represents a hypothetical area with minimal forest fragmentation, while the graphic on the right represents a hypothetical area with high forest fragmentation. Figure created by Jackson et al. (2006), used with permission of Oxford University Press.

Factors of disease prevalence in conjunction with landscape are present when examining the spatial distribution of Lyme disease. Frank et al. (1997) found that in a suburban area there was a correlation between the landscape and the presence of black-legged ticks in the nymph stage. In other words, the differences in land cover within a residential setting determined whether there was a high or low black-legged tick population. Another finding from Frank et al. (1997) concluded that suburban areas that had experienced some level of reforestation were correlated with areas of high abundance of black-legged ticks. Dister et al. (1997) made similar conclusions in a study where the researchers looked at how the landscape could impact the risk of Lyme disease prevalence. Their results showed that lawns next to, or nearby, woodlands were likely to show larger populations of black-legged ticks, therefore increasing the risk of Lyme disease transmission.

2.6 Conclusion

Previous studies have demonstrated the importance of both medical geography and landscape epidemiology concepts when researching the health risks of a region. More specifically, the spatial distribution and overall risk of Lyme disease emergence can be influenced by landscape epidemiology and forest fragmentation (Jackson et al., 2006). Transmitted by the black-legged tick, Lyme disease is a growing epidemiologic concern in the United States, and as discussed in prior studies, the way land use is allocated can impact the blood meal availability for black-legged ticks, the mobility of black-legged ticks, and the risk of transmission of the Lyme disease bacterium to humans (Brownstein et al., 2005, Finch et al., 2014, Hendricks & Mark-Carew, 2017, Jackson et al., 2006, Suekep et al., 2015). When looking at Virginia alone, there has been a considerable increase in the reported cases of human Lyme disease each year with the eastern shore, the northern, and the southwestern regions of Virginia having dense clusters of cases (Brinkerhoff et al., 2014, Lantos et al., 2015). With the geographic range and prevalence of Lyme disease increasing, it is important to consider how the land cover contributes to the emergence of Lyme disease (Brinkerhoff et al. 2014, Guerra et al., 2002, Kugeler et al., 2015, Lantos et al., 2015, Li et al., 2014).

References

Adrion, E. R., et al. (2015). "Health Care Costs, Utilization and Patterns of Care following Lyme Disease." PLOS ONE 10(2): e0116767.

Allan, B. F., et al. (2003). "Effect of Forest Fragmentation on Lyme Disease Risk." Conservation Biology 17(1): 267-272.

Brinkerhoff, R. J., et al. (2014). "Lyme disease, Virginia, USA, 2000-2011." Emerging infectious diseases 20(10): 1661.

Brownstein, J. S., et al. (2005). "Forest fragmentation predicts local scale heterogeneity of Lyme disease risk." Oecologia 146(3): 469-475.

Centers for Disease Control and Prevention (n.d.). *Lyme Disease (Borrelia burgdorferi) 2008 Case Definition*. Retrieved from https://wwwn.cdc.gov/nndss/conditions/lyme-disease/case-definition/2008/

Centers for Disease Control and Prevention. (2017a). *Data and Statistics*. Retrieved from https://www.cdc.gov/lyme/stats/index.html

Centers for Disease Control and Prevention. (2016a) *Lyme disease data tables: Reported cases of Lyme disease by state or locality, 2006-2016*. Retrieved from https://www.cdc.gov/lyme/stats/tables.html#modalIdString_CDCTable_0

Centers for Disease Control and Prevention. (2016b). *Signs and Symptoms of Untreated Lyme Disease*. Retrieved from https://www.cdc.gov/lyme/signs_symptoms/index.html

Centers for Disease Control and Prevention. (2019). *Transmission*. Retrieved from https://www.cdc.gov/lyme/transmission/index.html

Centers for Disease Control and Prevention. (2017b). *Why is CDC concerned about Lyme disease?* Retrieved from https://www.cdc.gov/lyme/why-is-cdc-concerned-about-lyme-disease.html

Cromley, E. K., et al. (1998). "Residential Setting as a Risk Factor for Lyme Disease in a Hyperendemic Region." American Journal of Epidemiology 147(5): 472-477.

Dister, S. W., et al. (1997). "Landscape Characterization of Peridomestic Risk for Lyme Disease Using Satellite Imagery." The American Journal of Tropical Medicine and Hygiene 57(6): 687-692.

Dymond, S. (2013). "Lyme Disease Emergence in Virginia: An Examination of the Demographic and Environmental Variables Correlated to the Spatial Pattern of Disease Incidence."

Finch, C., et al. (2014). "Integrated Assessment of Behavioral and Environmental Risk Factors for Lyme Disease Infection on Block Island, Rhode Island." PLOS ONE 9(1): e84758.

Frank, D. H., et al. (1998). "Landscape features associated with lyme disease risk in a suburban residential environment." Landscape Ecology 13(1): 27-36.

Guerra, M., et al. (2002). "Predicting the Risk of Lyme Disease: Habitat Suitability for Ixodes scapularis in the North Central United States." Emerging Infectious Diseases 8(3): 289-297.

Hendricks, B. and M. Mark-Carew (2017). "Using exploratory data analysis to identify and predict patterns of human Lyme disease case clustering within a multistate region, 2010-2014." Spatial and Spatio-Temporal Epidemiology 20: 35-43.

Jackson, L. E., et al. (2006). "Towards landscape design guidelines for reducing Lyme disease risk." International Journal of Epidemiology 35(2): 315-322.

Johnson, T. L., et al. (2017). "Prevalence and Diversity of Tick-Borne Pathogens in Nymphal Ixodes scapularis (Acari: Ixodidae) in Eastern National Parks." Journal of Medical Entomology 54(3): 742-751.

Kitron, U. (1998). "Landscape Ecology and Epidemiology of Vector-Borne Diseases: Tools for Spatial Analysis." Journal of Medical Entomology 35(4): 435-445.

Kugeler, K. J., et al. (2015). "Geographic Distribution and Expansion of Human Lyme Disease, United States." Emerging Infectious Diseases 21(8): 1455-1457.

Lambin, E. F., et al. (2010). "Pathogenic landscapes: Interactions between land, people, disease vectors, and their animal hosts." International Journal of Health Geographics 9(1): 54.

Lantos, P. M., et al. (2015). "Geographic Expansion of Lyme Disease in the Southeastern United States, 2000-2014." Open Forum Infectious Diseases 2(4).

Li, J., et al. (2014). "Spatial and Temporal Emergence Pattern of Lyme Disease in Virginia." The American Journal of Tropical Medicine and Hygiene 91(6): 1166-1172.

Mayer, J. D. (2000). "Geography, ecology and emerging infectious diseases." Social Science & Medicine: 937-952.

McGargal, K. and B. Marks (1995). "FRAGSTATS: spatial pattern analysis program for quantifying landscape structure." General Technical Report U.S. Department of Agriculture, Forest Service.

Meade, M. S. (2014). Medical Geography. The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society, John Wiley & Sons, Ltd.

Meentemeyer, R. K., et al. (2012). "Landscape Epidemiology of Emerging Infectious Diseases in Natural and Human-Altered Ecosystems." Annual Review of Phytopathology 50(1): 379-402.

Ostfeld, R. S. 2011. Lyme disease: the ecology of a complex system. New York: Oxford University Press.

Reisen, W. K. (2010). "Landscape Epidemiology of Vector-Borne Diseases." Annual Review of Entomology 55(1): 461-483.

Seukep, S. E., et al. (2015). "An Examination of the Demographic and Environmental Variables Correlated with Lyme Disease Emergence in Virginia." Ecohealth 12(4): 634-644.

Steere, A. C., Coburn, J., & Glickstein, L. (2004). “The emergence of Lyme disease.” Journal of Clinical Investigation, 113(8), 1093–1101. http://doi.org/10.1172/JCI200421681

CHAPTER 3:

THE EFFECTS OF LAND COVER ON THE SPATIAL DISTRIBUTION OF LYME DISEASE IN NORTHERN VIRGINIA SINCE 2005

Megan N. Stevenson[1] (*megsteve@vt.edu*), Korine N. Kolivras[1] (*korine@vt.edu*), Yang Shao[1] (y*shao@vt.edu*), Cassidy Rist[2] (*crist6@vt.edu*)

1. Virginia Tech Department of Geography, 115 Major Williams Hall, Blacksburg Virginia, 24060
2. Virginia Tech Department of Population Health Sciences, VA-MD College of Veterinary Medicine, 205 Duck Pond Drive, Blacksburg, Virginia 24061

Abstract

Human Lyme disease has been increasing in terms of case totals and geographic distribution since its identification in the 1970s. Consequently, Lyme disease is a growing public health concern and has been noted as one of the most common vector-borne illnesses within the United States. There is a wide range of prior studies that consider potential elements that have a role in the spread of Lyme disease, but few, if any, have analyzed the correlation between land use/land cover in northern Virginia and the incidence rate of human Lyme disease. As a result, this research aims to analyze the impact that land cover has on the spatial distribution of Lyme disease within the northern Virginia study area during the years 2005 to 2017.

The human Lyme disease case data for the study area was gathered by and received from the Virginia Department of Health (VDH). The National Land Cover Dataset (NLCD) 30 meter resolution images for the years 2006 and 2011 were gathered from the Multi-Resolution Land Characteristics Consortium (MRLC). The MRLC classifications were simplified to have four land cover types: i. Water, ii. Developed, iii. Forest, and iv. Herbaceous. ArcGIS Pro was used in addition to the Geospatial Modelling Environment software to collect and analyze information regarding land cover/land use elements. A Generalized Regression with Zero Inflation Gamma Distribution model was the chosen statistical test to determine if there was any correlation between the variables and human Lyme disease incidence within the study area.

As a result of this research, specific land use/land cover, such as; forest land cover type, herbaceous land cover type, and forest-herbaceous edge pairing were found to be correlated with human Lyme disease incidence rate in northern Virginia. In addition to land use/land cover, this research found a positive relationship between Lyme disease incidence rate and the number of occurrences of forest fragments less than 2ha, which suggests that Lyme disease in northern Virginia is also influenced by forest fragmentation.

Keywords:

Lyme disease, Medical Geography, Land Cover, GIS, ZI Gamma Distribution, Forest Fragmentation, Virginia

3.1 Introduction

The United States, specifically the eastern region, has experienced an emergence and increase in cases of human Lyme disease over the last several decades. Lyme disease was first identified in Lyme, Connecticut in the 1970s (Steere et al., 2004), and since then it has expanded in all directions (Brinkerhoff et al, 2014, Lantos et al., 2015). While Lyme disease was discovered in the 1970s, human cases of the disease only started to be reported and tracked in the United States by the Centers for Disease Control and Prevention (CDC) in the early 1990s (Kugeler et al., 2015).

Studies in the northeastern United States, the state of Virginia, and the midwestern United States have examined potential environmental drivers to explain the distribution and continued emergence of Lyme disease. The conclusions drawn from other studies suggest that land cover and land use influence the incidence rates of human Lyme disease in an area (e.g., Jackson et al. 2006a, Seukep et al., 2015), but no research has specifically examined environment-Lyme disease relationships in the northern Virginia area. The region is characterized by areas of dense urban and suburban development, without the large forested and agricultural swaths present in the remainder of the state. This study area shares characteristics with the region surrounding Lyme, Connecticut, yet differences may be present as the disease has emerged southward. We hypothesize that similar variables to those found in other studies examining Lyme-land cover relationships will be statistically significant here.

In an applied sense, the goal of this research is to identify characteristics tied to Lyme disease in an urban and suburban area that may be applicable to other regions with similar demographic and land cover characteristics. It is important for the overall public health in a community that land is developed in a manner that is cognizant of the risks that may potentially encourage Lyme disease spread (Jackson et al., 2006b). In northern Virginia, Lyme disease has a major presence and is fairly well known, but raising awareness of the disease and the land cover characteristics associated with cases is still of great importance as suburban development continues. Other areas outside of northern Virginia may be at risk of Lyme disease emergence and this research aims to contribute to decision-making among public health officials regarding the potential impact of land cover on the spread of Lyme disease. The two research questions that we sought to answer throughout this study were:

- What land cover characteristics are related to the emergence of Lyme disease in Northern Virginia from 2005 through 2017?
- Where should the Virginia Department of Health concentrate Lyme disease education and surveillance efforts when considering land cover types and the increasing spatial distribution of Lyme disease?

Drawn from reviewing relevant studies and methods, we hypothesized that there would be a correlation between the incidence rate of human Lyme disease and land cover variables in the northern Virginia study area. These hypotheses are plausible due to the similarities of characteristics found in northern Virginia and other Lyme endemic study areas from prior research.

3.2 Background

Along the east coast, Lyme disease is transmitted through the bite of *Ixodes scapularis*, more commonly known as the black-legged tick. The black-legged tick lifecycle lasts for about two years and the tick experiences three different stages of maturity during that time. In the beginning stages of the tick's life, the tick larvae tend to feed on small mammals, such as rodents and birds. It is during this first life stage that the black-legged ticks are likely to pick up the

Lyme disease bacterium, *Borrelia burgdorferi* (Allan et al., 2002). When looking at the east coast specifically, the white-footed mouse is the most competent reservoir of *Borrelia burgdorferi* and a common first blood meal for a black-legged tick. After the first blood meal, the ticks go almost a full year before feeding again. When the ticks reach the nymph stage they feed again, and it is during this stage that the black-legged ticks may choose to feed on humans and potentially transmit the bacterium if they were infected at the larval stage. While adult ticks also feed on humans, due to the small size of the nymphs, it is difficult to spot them and makes them a more likely contributor to the transmission of Lyme disease to humans than adult ticks (Allan et al., 2002, Brinkerhoff et al., 2014, Guerra et al., 2002, Johnson et al., 2017). It is also during these later stages of maturity that the ticks feed on white-tailed deer, which is not a reservoir for Lyme disease, but the deer do contribute to the spatial distribution of disease (Cromley et al., 1998, Dymond, 2013, Li et al., 2014). When an infected black-legged tick acquires a blood meal from a person, it is possible for the bacterium to then be transmitted, causing a human infection. Following a bite from an infected tick, the most common symptoms of human Lyme disease include headache, joint pain, fatigue, and other flu-like symptoms. About 70-80 percent of those infected with Lyme disease will develop a bullseye shaped rash, erythema migrans, around the site of the tick bite (CDC, 2016). Human Lyme disease can usually be effectively cured if a proper treatment regime is followed, which typically involves a prescription of doxycycline.

Land cover influences the spatial distribution of Lyme disease as it affects the distribution of the black-legged tick, reservoirs, and deer. A number of studies show the significant relationship between land use/land cover (LULC) types and their edges with Lyme disease and the spread of disease (e.g. Brownstein et al., 2005, Cromley et al., 1998, Jackson et al., 2006a, Li et al., 2014, Seukep et al., 2015). In general, research shows that regions with predominantly fragmented forest, forest-developed land cover edges or forest-herbaceous land cover edges are the best suited for Lyme disease emergence (Brownstein et al., 2005, Cromley et al., 1998, Jackson et al., 2006a). Jackson et al (2006a) suggest that the size and manner in which forests are fragmented plays a vital role in whether an area is at risk of Lyme disease emergence because forest fragmentation creates the ideal landscape configuration for two of the most common blood meals for a black-legged tick: the white-footed mouse and the white-tailed deer. White-footed mouse and white-tailed deer populations are more abundant in a fragmented forests due to the lack of predators, which generally do not thrive in a fragmented environment (Allan et al., 2003, Brownstein et al., 2005). In turn, the lack of predators allows for the black-legged tick population to thrive because of the plentiful opportunity for blood meals (Brownstein et al., 2005, Jackson et al., 2006a, Suekep et al., 2015). Consequently, there is an increased chance of black-legged ticks picking up *B. burgdorferi* in areas with a large presence of white-footed mice. Simon et al (2014) found that roughly 90% of larval ticks that receive a blood meal from an infected white-footed mouse end up infected themselves. Forest fragmentation is common in residential and suburban areas and such developments create the ideal size and layout of forest for all three populations (white-footed mouse, white-tailed deer, and black-legged tick) to thrive (Cromley et al., 1998, Dymond et al., 2013, Li et al., 2014, Mayer, 2000). These are also areas of high levels of human interaction with the environment since residential plots and walking trails are adjacent to or within the forest (Finch et al., 2014, Hendricks & Mark-Carew, 2017), increasing the likelihood of human infection.

Other studies have examined the correlation between LULC and Lyme disease incidence rate, but very few have focused on the effects of suburbanization on human Lyme disease prevalence specifically in the northern Virginia region. Since more and more areas are experiencing land cover change involving suburbanization, it is important to consider the impact that it has on the risk of disease, and in this case specifically, on the risk of Lyme disease emergence.

3.3 Study Area

As Lyme disease emerged southward, northern Virginia was the heaviest hit region in the state initially, mainly given the relatively high population. While several counties are notably more urbanized than the others, they are all considerably developed. With that said, each of the counties also contains farmland and undisturbed forested areas. Additionally, the region has developed land and forest-herbaceous edges that make them ideal places to analyze the distribution of Lyme disease, as prior research has shown that human Lyme disease incidence rates tend to be higher in areas with low to mid development and forest/herbaceous edges. Figure 3.1 outlines the counties and cities defined in this study as northern Virginia; they were chosen because of high incidence early in the epidemic and LULC conditions that are similar to other endemic areas.

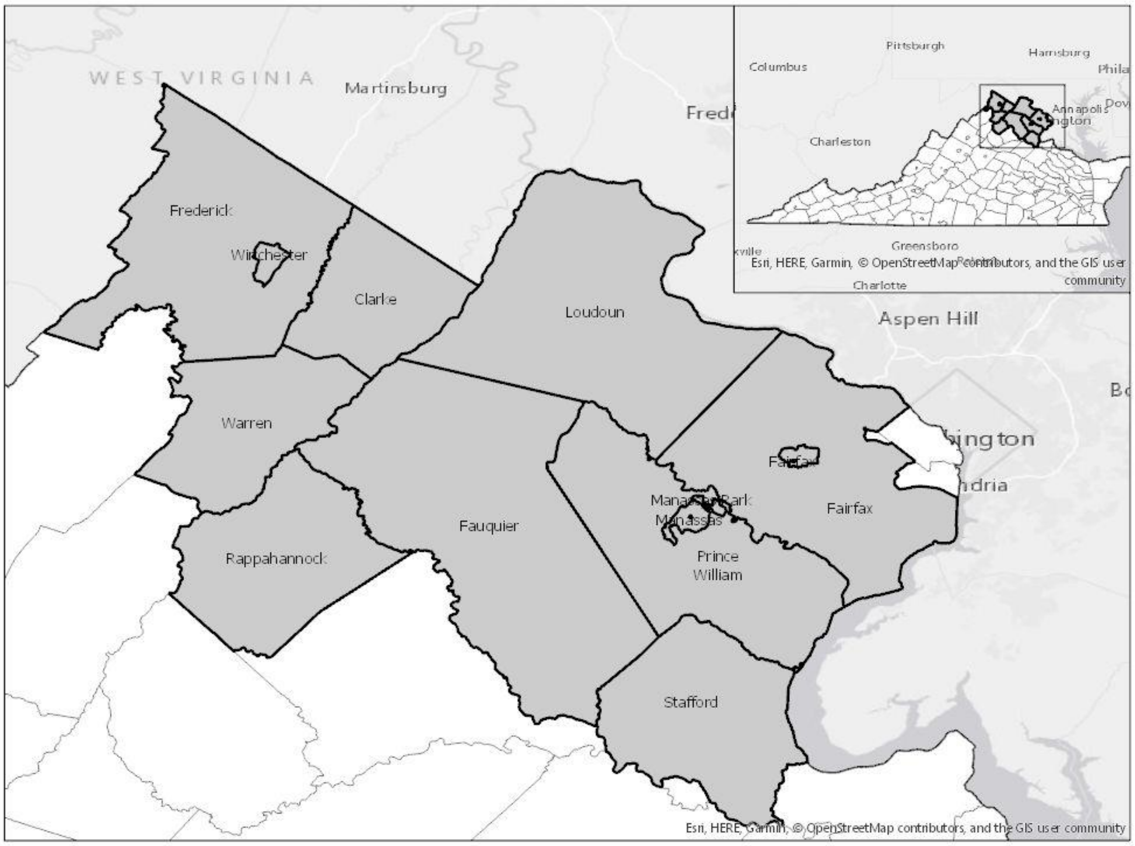

Figure 3.1: Defined study area includes: Clarke, Fairfax, Fauquier, Frederick, Loudoun, Prince William, Rappahannock, Stafford, and Warren counties, and Fairfax, Manassas, Manassas Park, and Winchester independent cities. County and state data layers from census.gov. Basemap from ESRI ArcPro software.

3.4 Data

Human Lyme disease case data was collected from the Virginia Department of Health (VDH) for 2005 to 2017. The data received from VDH included anonymous geo-coded point locations of the home addresses of reported Lyme disease cases, random case IDs, and the year of the reported case. The average number of human Lyme disease cases per year within the study area from 2005-2017 is approximately 446. During the same timeframe, the case total ranged from 170 in 2005 to 729 in 2010 (Figure 3.2). The cases of human Lyme disease are visualized on a map of the study area in Figure 3.3. Over the duration of the study period, the number of total cases fluctuates, as mentioned previously, but it remains constant that the largest cluster of cases appears in the northeastern portion of the study area (Figure 3.3). It is important to note that in 2008 the case definition for national reporting for Lyme disease changed (CDC, n.d.), which likely resulted in a decrease in the number of false negative cases and more accurate case totals (Brinkerhoff et al., 2014, Dymond, 2013).

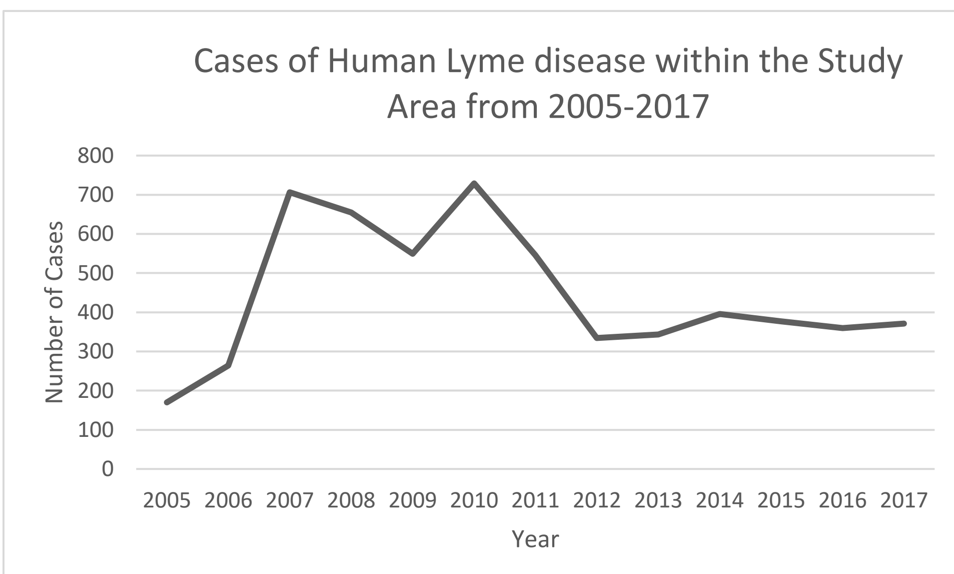

Figure 3.2: Total number of cases of human Lyme disease per year within the Northern Virginia study area from 2005-2017.

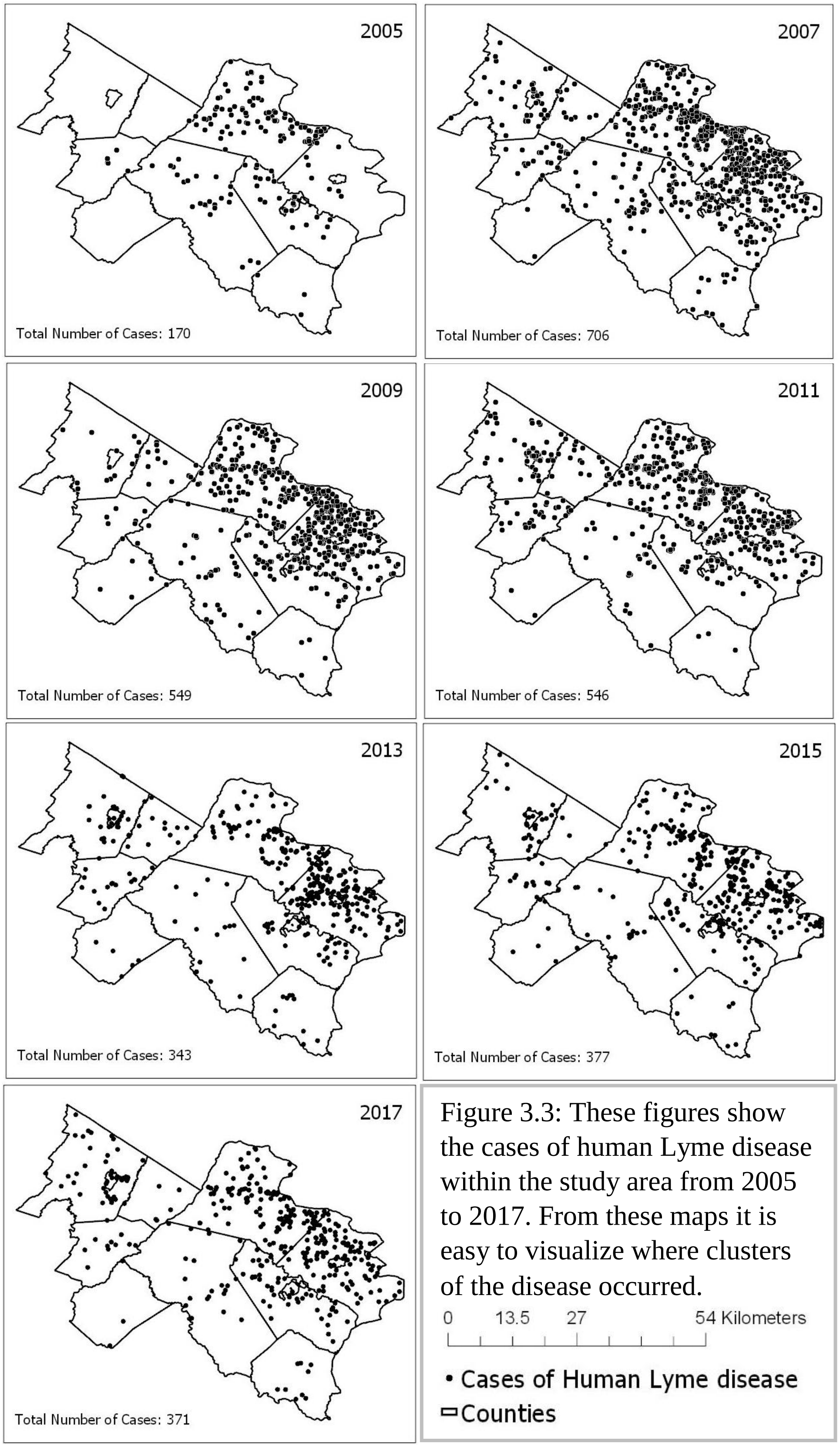

Figure 3.3: These figures show the cases of human Lyme disease within the study area from 2005 to 2017. From these maps it is easy to visualize where clusters of the disease occurred.

Figure 3.3: Total Number of Human Lyme disease cases in the Study Area from 2005-2017

National Land Cover Database (NLCD) images, at a spatial resolution of 30m, were acquired from the Multi-Resolution Land Characteristics Consortium (MRLC). NLCD images are Landsat-derived satellite images that have been classified, either manually or by machine learning methods, by the land use/land cover types present in the area that each image covers. There are 20 different land use types that the MRLC may classify in each image, including water, developed, barren, forest, and herbaceous types, and their more detailed subcategories. The images for this study were collected for 2006 and 2011; at the time of analysis, the NLCD imagery for 2016 was not yet available. For the analysis completed in this study, past land cover data was acceptable because current Lyme disease incidence rates are influenced by previous land cover trends that have established the transmission cycle. With that said, we chose to exclude the 2001 NLCD dataset due to the fact that our study timeframe began in 2005, in which case the 2006 NLCD dataset better represents the LULC characteristics of the year prior than the 2001 set would. Similarly, NLCD imagery from 2011 was applied to the examination of human case data beyond 2011 given that the 2016 dataset had not yet been released. Any new land cover trends that may have emerged and are visible in the 2016 land cover imagery set wouldn't be as influential or drastic as the older trends due to the presence of Lyme disease already having been established in the study area. With the known prevalence of Lyme disease in the region for many years prior to 2016, we can assume that the transmission cycle has been in motion since well before 2016.

Population totals for each census tract for each year within the study timeframe were derived from the US Census Bureau's American Fact Finder population datasets. The population data at the census tract level is only available for each decade. Therefore, it was necessary to extrapolate population estimates from the decennial Census data for each year of the study in between censuses. Population data used to create the maps and visualize the data (census tract layers, county layers, and state layer) were acquired from the US Census Bureau's Tiger/Line GIS datasets.

3.5 Methods

The first step was to visualize the data by displaying the reported cases of human Lyme disease on a map to identify regions that may have a high density of human Lyme disease cases within the study area (Figure 3.3 above). The cases of human Lyme disease were then aggregated to census tracts, and population totals per census tract per year were used to create an incidence rate of human Lyme disease per census tract for each year of the study (Figure 3.4). Maps were produced for each year of the study period to serve as visual aids for the change in human case totals over the years 2005 to 2017.

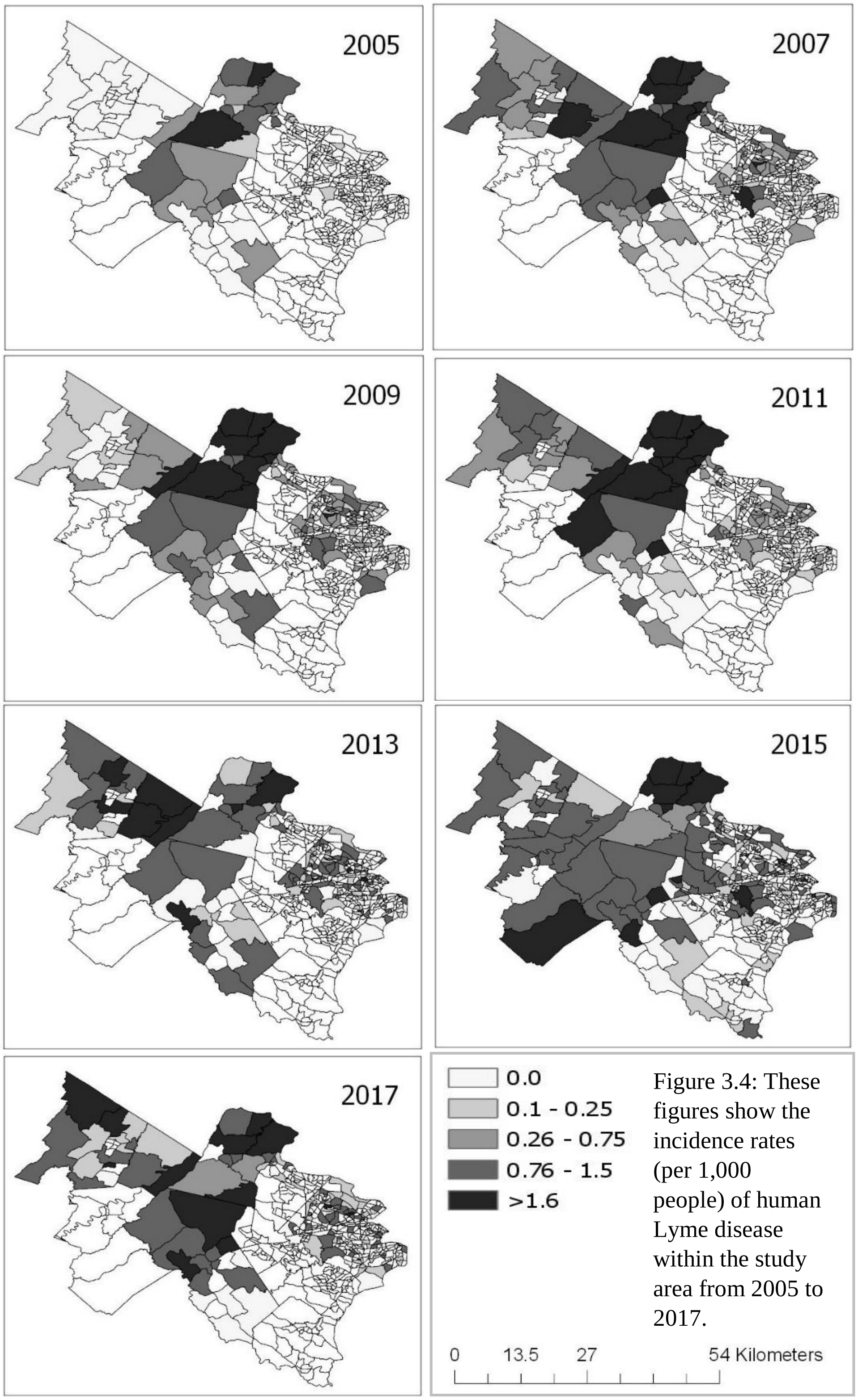

Figure 3.4: Incidence Rates (per 1,000 people) of Human Lyme disease in the Study Area from 2005-2017

Next, land cover datasets were processed. Land cover for both the 2006 and the 2011 images was re-classified using ArcGIS Pro into the following four broad land cover types: (i) water, (ii) developed, (iii) forest, (iv) herbaceous/agriculture. The forested classification type includes deciduous, evergreen, mixed forest, and woody wetlands. The herbaceous classification type includes developed open space, shrub/scrub, grassland/herbaceous, pasture/hay, cultivated crops, and emergent herbaceous wetlands. The developed open space was included in the herbaceous category because it is described as areas that are mostly grass within developed areas, such as residential lawns, golf courses, or other manmade landscapes (Multi-Resolution Characteristics Consortium, n.d.). Lastly, the developed classification type includes the low, medium, and high development classes from the NCLD original classifications as well as areas identified as barren since these areas would likely indicate land cleared for future development (Seukep et al., 2015). The water type was the only land cover classification category that did not require the aggregation and reclassification process because the dataset only had one type of water cover. From the reclassified NLCD images, we calculated the total hectares of forested land. We chose to then separate areas of forested land that were two hectares or less as forest fragments for use as a statistical variable based on the methods and findings of Suekep et al (2015).

When analyzing the relationship between reported cases of human Lyme disease and land cover, we used similar methods to those performed by Seukep et al. (2015). The Lyme disease incidence rate data were aggregated to their corresponding census tracts, and LULC within the census tracts were compared to the incidence rate. The NCLD images were also analyzed in terms of census tracts, and a land cover percentage was computed for each land cover type based on pixels (Seukep et al., 2015); however, we used Geospatial Modelling

Environment (GME) software instead of FRAGSTATS to conduct the land cover analysis. GME is the updated version of HawthsTools, and works with R statistical software as well as ArcGIS software to develop and run analyses. We first determined the percentage of each land cover type within each census tract. Then, given the previously identified importance of edges to the Lyme disease transmission cycle, we calculated edges between each land cover type within GME using the tool “Extract Edge” on each NLCD image. The purpose for extracting the edges of the four land cover types was to gain insight as to the association between land cover edge type pairings (water-developed, water-forest, water-herbaceous, developed-forest, developed-herbaceous, forest-herbaceous) and Lyme disease incidence rates. The output from the GME software was a GIS shapefile comprised of each possible land cover edge pairing and the attribute tables contained the total number of edges within each edge pairing category. From this output, the layers containing the edge pairings were spatially joined to the study area’s census tract layer. Information on the edges, such as the total edge length per edge pairing in each census tract and the number of occurrences of each land cover edge type, were calculated.

Finally, a generalized regression model with a Zero Inflation (ZI) Gamma distribution was used to determine the significance of the relationship between Lyme disease and the land cover variables within the census tracts described previously. This model was run using JMP statistical software and a maximum likelihood estimation method was used. The complete list of variables used in the statistical tests can be found in Table 3.1. A Zero Inflation Gamma distribution model was used due to the non-normal distribution of the incidence rate data and the imbalance of data caused by zeros throughout the dataset, given that some census tracts experienced no Lyme disease cases.

Table 3.1: Description of Variables for Statistical Analysis
Years for land cover-related data: 2006, 2011; years for incidence rate data: 2005-2017. Variable type for land cover-related data: explanatory, continuous; variable type for incidence rate data: response.

Variable Name
Percentage of water-developed edge length
Percentage of water-forest edge length
Percentage of water-herbaceous edge length
Percentage of developed-forest edge length
Percentage of developed-herbaceous edge length
Percentage of forest-herbaceous edge length
Frequency of forest fragments less than 2ha
Total hectares of forest fragments less than 2ha
Percentage of water
Percentage of developed land
Percentage of forest land
Percentage of herbaceous land
Incidence rate of human Lyme disease

Before running the ZI Gamma distribution model, Principal Component Analysis (PCA) was completed due to the multicollinearity present within the variables. Multicollinearity signifies that one or more of the variables were highly related to another and could therefore cause potential errors within the statistical test if left in their original state. PCA was run for each year of the NLCD data that was used in the study and the resulting components were calculated for each census tract. The variables included in each of the two PCA runs (for 2006 and 2011 NLCD layers) were percent of overall length for each land cover edge type, frequency (number of occurrences) of forest fragments less than 2ha, total hectares of forest fragments less than 2ha, and percent of each land cover type for their respective years of NLCD data (Table 3.1). From the PCA, three components were selected to be used in the ZI Gamma distribution model due to their significant impact on the data.

The ZI Gamma distribution model was run thirteen times in order to account for each year of case data. The years 2005-2010 were run through the ZI Gamma distribution model using the principal components for the 2006 NLCD imagery and dataset, while the years 2011-2017 were run using the principal components that were calculated with the 2011 NLCD imagery and corresponding dataset. In each test, the rate of Lyme disease incidence was represented as the response variable while the principal components were the explanatory variables (Table 3.1).

Since ZI Gamma distribution is a modified version of the normal generalized regression, the formula is different from that of the standard generalized regression model. Below is the formula for the generalized regression with ZI Gamma distribution model that was used for the analysis in JMP:

$$f(y|\mu, \sigma, \pi)=\left\{\begin{matrix} \pi, for\ y = 0 \\ (1-\pi)\dfrac{\exp(-y/\sigma)}{\Gamma[\mu/\sigma]\sigma^{\mu/\sigma}y^{1-\mu/\sigma}}, for\ y > 0 \end{matrix}\right\}$$

$$\mathrm{E}(\Upsilon) = \mu(1-\mu)$$

$$Var(\Upsilon) = \mu(1-\pi)(\sigma+\mu)-(1-\pi)^2\mu^2$$

3.6 Results

The results will be evaluated and discussed in terms of the two PCA groupings. First, we will share results for the set of principal components for the 2006 NLCD dataset the reported case years 2005-2010. When looking at the results of the model for each of these years, there are a few clear similarities. Overall, the principal component that suggested the greatest significance in relation to rates of human Lyme disease was principal component one. When looking at the variables in principal component one, the greatest weights were given to percent developed land, percent forested land, percent herbaceous land, percent of developed-herbaceous edge, percent of developed-forest edge, and percent of forest-herbaceous edge. Individually speaking, percent forested land, percent herbaceous land, and percent forest-herbaceous edge each had a positive relationship with Lyme disease incidence rates, while percent developed land, percent developed-herbaceous edge, and percent developed-forest edge each had a negative relationship with Lyme disease incidence rates. Principal component one is statistically significantly related to incidence rate in 2007 through 2010, but the years 2005 and 2006 do not show significance for any of the principal components (Table 3.2).

Table 3.2: Results of Statistical Analysis for Principal Component 1, Years 2005-2010

Year	**Estimate**	**Standard Error**	**Prob > ChiSquare**	**Dispersion Prob > ChiSquare**	**Zero Inflation Prob > ChiSquare**
2005	0.0461391	0.0573771	0.4213	**0.0021***	**<0.0001***
2006	0.0674161	0.0521952	0.1965	**<0.0001***	**<0.0001***
2007	0.1311626	0.0377439	**0.0005***	**<0.0001***	**<0.0001***
2008	0.136837	0.037109	**0.0002***	**<0.0001***	**<0.0001***
2009	0.1053948	0.0327447	**0.0013***	**<0.0001***	**<0.0001***
2010	0.1342437	0.0279163	**0.0001***	**<0.0001***	**<0.0001***

* - statistically significant (α=0.05) If the value is in bold that symbolizes that the p-value is less than 0.01.

The second set of principal components was developed for the 2011 NLCD dataset and was used when running the ZI Gamma regression model for the case years 2011-2017. Similarly to the 2006 PCA grouping, the years in the 2011 PCA grouping showed a significant correlation between Lyme disease incidence rates and principal component one. Just as with the 2006 principal component one, the variables in principal component one for the 2011 NLCD data set were percent developed land, percent forested land, percent herbaceous land, percent developed-herbaceous edge, percent developed-forest edge, and percent forest-herbaceous edge. As we saw with the 2006 dataset, when looking at the individual variable level for the principal component the percent forested land, percent herbaceous land, and percent forest-herbaceous edge each had a positive relationship with Lyme disease incidence rates, while percent developed land, percent developed-herbaceous edge, and percent developed-forest edge each had a negative relationship with Lyme disease incidence rates. Similar to what we saw with the 2006 PCA grouping, every year in the 2011 set except for one year (2013) showed a significant correlation between Lyme disease incidence rate and the variables within principal component one (Table 3.3). The year 2013 showed a significant correlation between Lyme disease incidence rate and principal component three, which places the greatest weights on the following variables: percent water cover, percent herbaceous land, percent water-developed edge, and percent water-herbaceous edge. When principal component three is broken down to the individual variable level, percent water cover, percent herbaceous land, percent water-developed edge, and percent water-herbaceous edge all show a positive relationship to Lyme disease incidence. The p-value for principal component three in 2013 is 0.0316, which is above the p-values for principal component one for the remaining years. The p-values of 2011, 2012, and 2014-2017 are all below the 0.01 threshold, which suggests that the significance of the correlation between principal component one and Lyme disease incidence rate in those years is greater than that of principal component three in the year 2013.

Table 3.3: Results of Statistical Analysis for Principal Component 1, Years 2011-2017

Year	**Estimate**	**Standard Error**	**Prob > ChiSquare**	**Dispersion Prob > ChiSquare**	**Zero Inflation Prob > ChiSquare**
2011	0.1303276	0.0426317	**0.0022***	**<0.0001***	**<0.0001***
2012	0.1180992	0.0324729	**0.0003***	**<0.0001***	**<0.0001***
2013	0.0229798	0.0273671	0.4011	**<0.0001***	**<0.0001***
2014	0.0985859	0.0370964	**0.0079***	**<0.0001***	**<0.0001***
2015	0.0831594	0.0235258	**0.0004***	**<0.0001***	**<0.0001***
2016	0.0949145	0.0386931	**0.0142***	**<0.0001***	**<0.0001***
2017	0.1206976	0.037233	**0.0012***	**<0.0001***	**<0.0001***
* - statistically significant (α=0.05) If the value is in bold that symbolizes that the p-value is less than 0.01.					

Beyond the standard results of the statistical test, there are other output variables to consider since the normal gamma regression model was altered to resemble a zero inflation model. When analyzing the overall results for each year and each PCA grouping, it was important to also look at the p-values of the zero inflation variable and the dispersion variable. Both of these variables have significant p-values for all thirteen years of the data, which suggests that the high number of zeros in the Lyme disease incidence rate dataset would have greatly influenced the results of the statistical test if it had not been considered prior to running the model. The same could be said about the abnormal dispersion of the data.

Table 3.4 shows the average percent change of each of the variables found in principal component one between the 2006 land cover image and the 2011 land cover image. Though the changes are relatively minimal, they can impact the Lyme disease transmission cycle and, therefore, the overall case totals in the study area throughout the years. The percentage of developed-forest edge length and percentage of forest-herbaceous edge length both decreased from 2006 to 2011, while the percentage of developed-herbaceous edge length increased. Lastly, there was a decrease in both percentage of land classified as forest (-1.42%) and percentage of land classified as herbaceous (-0.19%) from 2006 to 2011. During this same time period, there was a +1.67% increase in the percentage of land classified as developed. This percent change of land use represents how the landscape in the study area was altered over the years in favor of developing more urban space and smaller forest fragments, all while maintaining relatively stable levels of herbaceous land cover.

Table 3.4: Principal Component One Variables Percent Change from 2006 to 2011

Variable Name	**Average Percentage in 2006**	**Average Percentage in 2011**	**Percent Change**
Percentage developed-forest edge length	8.37%	8.13%	-0.23%
Percentage developed-herbaceous edge length	55.00%	55.62%	+0.62%
Percentage forest-herbaceous edge length	35.13%	34.24%	-0.89%
Percentage developed land	39.08%	40.75%	+1.67%
Percentage forest land	28.83%	27.41%	-1.42%

Percentage herbaceous land	30.08%	29.89%	-0.19%

3.7 Discussion

The purpose of this study was to determine if correlations exist between land cover types and edges, and human Lyme disease incidence rate in northern Virginia. In particular, we sought to examine if relationships identified in endemic Lyme disease areas north of the study area (e.g. New England) were also identified as the disease emerged southward into northern Virginia. The case data and incidence rates were visualized using maps created in ArcGIS Pro that included overlays of land cover types, and the results of the statistical tests offer insight as to which land cover types are associated with higher incidence rates of human Lyme disease.

This research confirms our hypothesis that land use/land cover is a key contributor to the spread of Lyme disease. The research findings also suggest that forest fragmentation and landscape configuration with respect to edges are also important factors in contributing to human Lyme disease incidence rate. This claim is supported by the presence of the positive relationship between Lyme disease incidence and the frequency of forest fragments <2ha. The results gathered from this study support the findings of similar studies that have been conducted on the impacts of land cover and Lyme disease occurrence (e.g., Jackson et al., 2006a, Seukep et al., 2015). Prior studies have similarly concluded that smaller forest fragments are positively related to human Lyme disease incidence due to the creation of more edges between land cover types, specifically when the resulting edges are forest-herbaceous (e.g., Brownstein et al., 2005, Jackson et al., 2006a, Seukep et al., 2015). Based on previous studies, we hypothesized that the northern Virginia region would see a correlation between land cover and human Lyme disease incidence rate given the presence of commonalities between northern Virginia and other Lyme endemic regions that have been studied in the past, and our results allow us to confirm that hypothesis.

Examining the case totals from 2005-2017, we can see a fluctuation in the number of cases (Figure 3.2). In 2005 the total number of human cases of Lyme disease within the study area was 170, but by 2007 the yearly total of cases had risen to 706. The total number of annual cases of Lyme disease began to decrease in 2008 and continued to gradually decline in the years following 2008. Since 2012 the total number of cases per year has remained fairly stable hovering in the mid to upper three hundred cases range. This fluctuation and eventual stabilization could be partly attributed to the alteration of the land that either supports or constrains the disease transmission cycle.

When looking at the percentage of developed-forest edge length and percentage of forest-herbaceous edge length for the years 2006 and 2011, both percentages decreased. Conversely, the percentage of developed-herbaceous edge length increased during the same time period. It is possible that the forests in the study area were further fragmented within the 2006-2011 timeframe to allow for more development. An increase in development could also attribute for the increase in percentage of developed-herbaceous edge length, which is a contributing factor to the spatial distribution of Lyme disease as seen in the statistical results.

Statistical test results suggest a correlation between Lyme disease incidence rates and forest and herbaceous edge types, along with overall forest and herbaceous land cover types. These results are logical when considering results of previous studies that show the success of black-legged tick populations and the increased likelihood of human interaction with the environment in regions with these land cover characteristics and configurations. Both developed-forest edge and developed-herbaceous edge showed a negative relationship to human Lyme disease incidence rates even though previous studies have shown that there could be a correlation between Lyme disease incidence rate and suburban sprawl (Brownstein et al., 2005, Cromley et al., 1998, Dymond et al., 2013, Li et al., 2014, Mayer, 2000). The negative trends between the developed-forest and developed-herbaceous edge types and Lyme disease incidence rate could be due to the density of developed land within the developed land classification. As mentioned earlier, the original developed land classification levels (low, medium, high) were aggregated to create a single developed classification. The aggregation of developed land classification types could have resulted in an oversimplification of the developed land cover type, which in turn could suggest higher levels of density of development within the census tracts than would be seen if the original classifications had been used. The way in which the developed-forest or developed-herbaceous edges are designed could also impact the dispersion of human Lyme disease cases, which relates to Jackson et al.'s (2006a and 2006b) point that the particular manner in which land is fragmented and developed can affect the spread of Lyme disease. With that said, our study area may contain suburban areas that are not conducive to the Lyme disease transmission cycle due to the way they were developed and the quantity/dispersion of green space. Three years' results differed from other years: 2005, 2006, and 2013. A correlation was not present between land cover and Lyme disease incidence rate in 2005 and 2006, while in the year 2013, a significant relationship was identified between Lyme disease incidence rate and percent of water cover, percent of herbaceous land, percent of water-developed edge, and percent of water-herbaceous edge (principal component three for NLCD 2011 set). These unexpected findings could have resulted due to a number of issues, such as case reporting, NLCD image classification errors, and statistical uncertainty.

When contemplating how the land could be developed in a way to reduce the risk of human Lyme disease, it's important to take forest fragment size into consideration. As seen in prior research and supported by findings in this study, forest fragments that are two hectares or less show a correlation with Lyme disease incidence rates (Suekep et al., 2015). With that in mind, it would be wise for development plans to maintain forest patches greater than two hectares. By developing areas, specifically residential communities, with fewer forest edges and forest fragments greater than two hectares, it allows for a reduced risk of Lyme disease in the area while also creating a community within close proximity to nature and walking trails, which is attractive to suburban home buyers.

This research contains limitations due to the nature and scope of the data and how data were collected. Working with reported cases of human Lyme disease can present a few concerns. First, the locations given by VDH for reported cases represents where a patient lives, but that may not always also be the point of infection. Second, underreporting is a known issue for Lyme disease (CDC, 2018), and therefore we are analyzing only those cases that were diagnosed and reported. Lastly, the recognition of the change in case definition in 2008 and the effects that had on case totals needs to be acknowledged, along with the overall accuracy of the Lyme disease blood test. Prior to 2008, the case definition allowed for more cases to be diagnosed as Lyme disease, whereas now there are stricter rules and determinants for the disease.

Another limitation is the potential for misclassification of the NLCD images. It is possible that the land cover identified in each of the images is not completely accurate. In the case of misclassification, conclusions could be drawn regarding which land cover types support the Lyme disease transmission cycle that may not be accurate. For the scope and scale of this study, 30 meter resolution NLCD images are the best option for land cover data.

The results of this research represent a level of scale dependency, and therefore, the scale of the data used in this study could be a limitation. We chose to analyze the data at the census tract level, but the results of the study could vary if an even larger scale was used, such as block groups. The use of larger scale data would allow for a more in depth analysis of the study area and Lyme disease incidence rate trends. Future research could be improved by including larger scale data in the analysis to provide a better understanding of the disease spread patterns in a region.

There are a multitude of factors that contribute to Lyme disease emergence, many of which were not included in this study as we chose to focus on land cover characteristics. Due to the research's focus on the effects of land use/land cover, additional environmental and locational variables were not taken into account, which could be considered a limitation to this research and the resulting conclusions. This research did not include elevation, climate, and soil type on tick prevalence in the study area, or the potential role of human behavior in potential tick exposure, as we chose to focus on land use/land cover. As a result, this study could be improved with the addition of environmental or demographic data in the analysis. Future research efforts could aim to evaluate a more comprehensive list of factors that contribute to Lyme disease prevalence and their impacts on human Lyme disease incidence rate.

3.8 Conclusion

This chapter evaluated land cover-Lyme disease relationships within northern Virginia to determine if there is a correlation between forest fragmentation, specific land cover types and configurations, suburbanization, and cases of human Lyme disease, as has been demonstrated in other Lyme endemic regions in the United States. Few studies have focused specifically on northern Virginia when considering the impacts of land cover change and suburban sprawl on Lyme disease, despite a high number of cases in the region.

The United States has seen an increase in human cases of Lyme disease and an expansion of the overall spatial distribution of the disease over the last few decades. As a result, Lyme disease is now considered to be the most common vector-borne illness in the US (CDC, 2017b and Rosenberg et al., 2018). The CDC (2017) estimates that there are nearly 300,000 cases of human Lyme disease every year with many cases going unreported due to lack of diagnoses or failure of physicians to report. The spatial distribution of Lyme disease in the United States continues to expand, which makes understanding the changing endemic zone of critical importance. This research analyzed the relationship between Lyme disease at the regional scale to contribute to our understanding of which land cover elements have the greatest effect on the disease.

The results of this study contribute to the existing knowledge of Lyme disease and further enforce the evidence that its spatial distribution is impacted by land cover. We found that Lyme disease in northern Virginia is influenced by land cover fragmentation, certain land cover types, and specific edges. This study supports the claim that the number of human Lyme disease cases will be greater in areas that have fragmented forests and forest-herbaceous edges in residential communities due to the positive relationship between human Lyme disease incidence and forest-herbaceous edges as well as frequency of fragmented forests that are two hectares or less.

In an applied sense, the findings of this research can be applicable to other regions of similar urban design so that proper precautions can be taken both in land development and in ensuring public health safety through education efforts. Ideally, regions with similar demographics, environmental characteristics, and land use elements would be more aware of the risks of land use change and suburban sprawl in relation to Lyme disease and understand the need to be proactive as a result of this study. The findings from this study contribute to the growing knowledge base of Lyme disease-land cover links and should inspire further research to be conducted to work towards a greater understanding of the disease and how to best slow its expansion.

References

Allan, B. F., et al. (2003). "Effect of Forest Fragmentation on Lyme Disease Risk." Conservation Biology 17(1): 267-272.

Beyer, H., Spatial Ecology LLC. (2012). Geospatial Modelling Environment [Computer software]. Retrieved from http://www.spatialecology.com/gme/gmedownload.htm

Brinkerhoff, R. J., et al. (2014). "Lyme disease, Virginia, USA, 2000-2011." Emerging infectious diseases 20(10): 1661.

Brownstein, J. S., et al. (2005). "Forest fragmentation predicts local scale heterogeneity of Lyme disease risk." Oecologia 146(3): 469-475.

Centers for Disease Control and Prevention (n.d.). *Lyme Disease (Borrelia burgdorferi) 2008 Case Definition*. Retrieved from https://wwwn.cdc.gov/nndss/conditions/lyme-disease/case-definition/2008/

Centers for Disease Control and Prevention. (2017a). *Data and Statistics*. Retrieved from https://www.cdc.gov/lyme/stats/index.html

Center for Disease Control and Prevention. (2018). *How many people get Lyme Disease?* Retrieved from https://www.cdc.gov/lyme/stats/humancases.html

Centers for Disease Control and Prevention. (2016). *Signs and Symptoms of Untreated Lyme Disease*. Retrieved from https://www.cdc.gov/lyme/signs_symptoms/index.html

Centers for Disease Control and Prevention. (2017b). *Why is CDC concerned about Lyme disease?* Retrieved from https://www.cdc.gov/lyme/why-is-cdc-concerned-about-lyme-disease.html

Cromley, E. K., et al. (1998). "Residential Setting as a Risk Factor for Lyme Disease in a Hyperendemic Region." American Journal of Epidemiology 147(5): 472-477.

Dymond, S. (2013). "Lyme Disease Emergence in Virginia: An Examination of the Demographic and Environmental Variables Correlated to the Spatial Pattern of Disease Incidence."

Esri. (2018). ArcGIS Pro: Release 2.2.1 [Computer software]. Retrieved from https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview

Finch, C., et al. (2014). "Integrated Assessment of Behavioral and Environmental Risk Factors for Lyme Disease Infection on Block Island, Rhode Island." PLOS ONE 9(1): e84758.

Guerra, M., et al. (2002). "Predicting the Risk of Lyme Disease: Habitat Suitability for Ixodes scapularis in the North Central United States." Emerging Infectious Diseases 8(3): 289-297.

Hendricks, B. and M. Mark-Carew (2017). "Using exploratory data analysis to identify and predict patterns of human Lyme disease case clustering within a multistate region, 2010-2014." Spatial and Spatio-Temporal Epidemiology 20: 35-43.

Jackson, L. E., et al. (2006a). "Towards landscape design guidelines for reducing Lyme disease risk." International Journal of Epidemiology 35(2): 315-322.

Jackson L.E., J.F. Levine, E.D. Hilbourn (2006b). "A comparison of analysis units for associating Lyme disease with forest-edge habitat." Community Ecology: 189–197

Johnson, T. L., et al. (2017). "Prevalence and Diversity of Tick-Borne Pathogens in Nymphal Ixodes scapularis (Acari: Ixodidae) in Eastern National Parks." Journal of Medical Entomology 54(3): 742-751.

Kugeler, K. J., et al. (2015). "Geographic Distribution and Expansion of Human Lyme Disease, United States." Emerging Infectious Diseases 21(8): 1455-1457.

Lantos, P. M., et al. (2015). "Geographic Expansion of Lyme Disease in the Southeastern United States, 2000-2014." Open Forum Infectious Diseases 2(4).

Li, J., et al. (2014). "Spatial and Temporal Emergence Pattern of Lyme Disease in Virginia." The American Journal of Tropical Medicine and Hygiene 91(6): 1166-1172.

Mayer, J. D. (2000). "Geography, ecology and emerging infectious diseases." Social Science & Medicine: 937-952.

Multi-Resolution Land Characteristics Consortium (MRLC). (2019b). NLCD 2006 Land Cover Conterminous United States. Retrieved from https://www.mrlc.gov/data?f%5B0%5D=category%3Aland%20cover&f%5B1%5D=year%3A2006

Multi-Resolution Land Characteristics Consortium (MRLC). (2019b). NLCD 2011 Land Cover Conterminous United States. Retrieved from https://www.mrlc.gov/data?f%5B0%5D=category%3Aland%20cover&f%5B1%5D=year%3A2011

Multi-Resolution Land Characteristics Consortium (MRLC). (n.d.). National Land Cover Database 2011 (NLCD2011) Legend. Retrieved from https://www.mrlc.gov/data/legends/national-land-cover-database-2011-nlcd2011-legend

Rosenberg, R., et al. (2018). "Vital Signs: Trends in Reported Vectorborne Disease Cases - United States and Territories, 2004–2016." Retrieved from https://www.cdc.gov/mmwr/volumes/67/wr/pdfs/mm6717e1-H.pdf

SAS Institute. (2018). JMP Pro 14 [Computer software]. Retrieved from https://www.jmp.com/en_us/software/predictive-analytics-software.html

Seukep, S. E., et al. (2015). "An Examination of the Demographic and Environmental Variables Correlated with Lyme Disease Emergence in Virginia." Ecohealth 12(4): 634-644.

Simon, J. A., et al. (2014). "Climate change and habitat fragmentation drive the occurrence of Borrelia burgdorferi, the agent of Lyme disease, at the northeastern limit of its distribution." Evolutionary Applications 7(7): 750–764.

Steere, A. C., Coburn, J., & Glickstein, L. (2004). "The emergence of Lyme disease." Journal of Clinical Investigation, 113(8), 1093–1101. http://doi.org/10.1172/JCI200421681

United States Census Bureau, Geography Division. (2017). 2017 TIGER/Line Shapefiles. Retrieved from https://www.census.gov/cgi-bin/geo/shapefiles/index.php